#6

SEQUENCE LISTING

(1) GENERAL INFORMATION:

   (i) APPLICANT: Robert F.M. Van Gorcom
                 Willem Van Hartingsveldt
                 Petrus A. Van Paridon
                 Annemarie E. Veenstra
                 Rudolf G.M. Luttin
                 Gerardus Selten

   (ii) TITLE OF INVENTION: Cloning and Expression of Microbial
       Phytase

  (iii) NUMBER OF SEQUENCES: 52

   (iv) CORRESPONDENCE ADDRESS:
       (A) ADDRESSEE: Morrison & Foerster
       (B) STREET: 545 Middlefield Road, Suite 200
       (C) CITY: Menlo Park
       (D) STATE: California
       (E) COUNTRY: USA
       (F) ZIP: 94025-3471

   (v) COMPUTER READABLE FORM:
       (A) MEDIUM TYPE: Floppy disk
       (B) COMPUTER: IBM PC compatible
       (C) OPERATING SYSTEM: PC-DOS/MS-DOS
       (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
          (EPO)

   (vi) CURRENT APPLICATION DATA:
       (A) APPLICATION NUMBER: 07/688,578
       (B) FILING DATE: 24-MAY-1991
       (C) CLASSIFICATION:

 (viii) ATTORNEY/AGENT INFORMATION:
       (A) NAME: Murashige, Kate H.
       (B) REGISTRATION NO: 29,959
       (C) REFERENCE/DOCKET NUMBER: 24615-20026.00

   (ix) TELECOMMUNICATION INFORMATION:
       (A) TELEPHONE: 415-327-7250
       (B) FACSIMILE: 415-327-2951

(2) INFORMATION FOR SEQ ID NO:1:

   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 9 amino acids
       (B) TYPE: amino acid
       (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

  (iii) HYPOTHETICAL: NO

1

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Gln Ser Ser Xaa Asp Thr Val Asp Gln
1   5

(2) INFORMATION FOR SEQ ID NO:2:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 35 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (iii) HYPOTHETICAL: NO

    (v) FRAGMENT TYPE: N-terminal

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Ala Ser Xaa Xaa Gln Ser Ser Xaa Asp Thr Val Asp Gln Gly Tyr Gln
1   5   10   15

Arg Phe Ser Glu Thr Ser His Leu Arg Xaa Gln Tyr Ala Pro Phe Phe
            20   25   30

Asp Leu Ala
        35

(2) INFORMATION FOR SEQ ID NO:3:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 10 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (iii) HYPOTHETICAL: NO

    (v) FRAGMENT TYPE: N-terminal

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Val Val Asp Glu Arg Phe Pro Tyr Thr Gly
1   5   10

(2) INFORMATION FOR SEQ ID NO:4:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 20 amino acids

```
            (B)  TYPE: amino acid
            (D)  TOPOLOGY: linear

     (ii)  MOLECULE TYPE: peptide

    (iii)  HYPOTHETICAL: NO

      (v)  FRAGMENT TYPE: internal

     (xi)  SEQUENCE DESCRIPTION: SEQ ID NO:4:

     Gln Xaa Gln Ala Glu Gln Glu Pro Leu Val Arg Val Leu Val Asn Asp
     1               5                   10                  15

     Arg Val Val Pro
                 20


(2) INFORMATION FOR SEQ ID NO:5:

      (i)  SEQUENCE CHARACTERISTICS:
            (A)  LENGTH: 23 amino acids
            (B)  TYPE: amino acid
            (D)  TOPOLOGY: linear

     (ii)  MOLECULE TYPE: peptide

    (iii)  HYPOTHETICAL: NO

      (v)  FRAGMENT TYPE: internal

     (xi)  SEQUENCE DESCRIPTION: SEQ ID NO:5:

     Xaa Ser Phe Asp Thr Ile Ser Thr Ser Thr Val Asp Thr Lys Leu Ser
     1               5                   10                  15

     Pro Phe Cys Asp Leu Phe Thr
                 20


(2) INFORMATION FOR SEQ ID NO:6:

      (i)  SEQUENCE CHARACTERISTICS:
            (A)  LENGTH: 16 amino acids
            (B)  TYPE: amino acid
            (D)  TOPOLOGY: linear

     (ii)  MOLECULE TYPE: peptide

    (iii)  HYPOTHETICAL: NO

      (v)  FRAGMENT TYPE: N-terminal

     (xi)  SEQUENCE DESCRIPTION: SEQ ID NO:6:

     Leu Ala Val Pro Ala Ser Arg Asn Gln Ser Ser Gly Asp Thr Val Asp
     1               5                   10                  15
```

3

(2) INFORMATION FOR SEQ ID NO:7:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 19 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

   (iii) HYPOTHETICAL: NO

    (v) FRAGMENT TYPE: internal

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Met Gln Cys Gln Ala Glu Gln Glu Pro Leu Val Arg Val Leu Val
1   5   10   15

Asn Asp Arg

(2) INFORMATION FOR SEQ ID NO:8:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

   (iii) HYPOTHETICAL: NO

    (v) FRAGMENT TYPE: internal

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Ala Ser Ser Ala Glu Lys Gly Tyr Asp Leu Val Val
1   5   10

(2) INFORMATION FOR SEQ ID NO:9:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

   (iii) HYPOTHETICAL: NO

    (v) FRAGMENT TYPE: internal

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
Val Val Asp Xaa Arg Phe Pro Tyr Thr Gly Xaa Ala
1             5                   10
```

(2) INFORMATION FOR SEQ ID NO:10:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 116 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA

    (iii) HYPOTHETICAL: YES

    (vi) ORIGINAL SOURCE:
        Phytase N-terminus reverse
            translation

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

YTNGCNGTNC CNGCNWSNMG NAAYCARWSN WSNGGNGAYA CNGTNGAYCA RGGNTAYCAR   60

MGNTTWWWSA RACNWSNCAW YTNMGNGGNC ARTAYGCNCC NTTYTTYGAY YTNGCN   116

(2) INFORMATION FOR SEQ ID NO:11:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 66 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA

    (iii) HYPOTHETICAL: YES

    (vi) ORIGINAL SOURCE:
        internal fragment A (Phytase)
            reverse translation

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CARNNNCARG CRGANCARGA RCCRYTNGTN HSNGTNYTNG TNRAYVVNVK NGTNCCNCCN   60

ATGGGN   66

(2) INFORMATION FOR SEQ ID NO:12:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 99 base pairs
        (B) TYPE: nucleic acid

            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA

   (iii) HYPOTHETICAL: YES

    (vi) ORIGINAL SOURCE:
            internal fragment B (Phytase)
                    reverse translation

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TGGWSNTTYG AYACNATHWS NACNWSNACN GTNGAYACNA ARYTNWSNCC NTTYTCYGAY   60

YTNTTYACNA CNGAYGARTG YATHAMNTAY VGNTAYYTN   99

(2) INFORMATION FOR SEQ ID NO:13:

    (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 69 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA

   (iii) HYPOTHETICAL: YES

    (vi) ORIGINAL SOURCE:
            alkaline phosphatase reverse
                    translation

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TTYWSNTAYG GNGCNGCNAT HCCNCARWSN ACNCARGARA ARCARTTYWS NCARGARTTY   60

MGNGAYGGN   69

(2) INFORMATION FOR SEQ ID NO:14:

    (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 48 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

   (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
            AB1024

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTGGTCGACG GTGTCGCCGC TGCTCTGGTT GCGGCTGGCG GGGACGGC   48

(2) INFORMATION FOR SEQ ID NO:15:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 18 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

    (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
        AB1065

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CTGRTCCACG GTGTCGCC   18

(2) INFORMATION FOR SEQ ID NO:16:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 18 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

    (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
        AB1066

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTGRTCGACG GTGTCGCC   18

(2) INFORMATION FOR SEQ ID NO:17:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 18 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

    (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
          AB1067

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CTGRTCCACA GTGTCGCC   18

(2) INFORMATION FOR SEQ ID NO:18:

    (i)   SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 18 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

    (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
          AB1069

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTGRTCCACG GTATCGCC   18

(2) INFORMATION FOR SEQ ID NO:19:

    (i)   SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 18 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

    (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
          AB1069

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CTGGTCCACG GTGTCACC   18

(2) INFORMATION FOR SEQ ID NO:20:

    (i)   SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 18 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
     AB1070

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CTGATCGACA GTATCACC   18

(2) INFORMATION FOR SEQ ID NO:21:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 18 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

    (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
         AB1226

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CTGGTARCCC TGRTCSAC   18

(2) INFORMATION FOR SEQ ID NO:22:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 18 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

    (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
         AB1227

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

YTGRTADCCY TGRTCVAC   18

(2) INFORMATION FOR SEQ ID NO:23:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 24 base pairs

9

            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (synthetic)

       (iii) HYPOTHETICAL: NO

        (vi) ORIGINAL SOURCE:
              AB1298

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

YTGRTASCCK TGRTCSACSG TRTC   24

(2) INFORMATION FOR SEQ ID NO:24:

         (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 26 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (synthetic)

       (iii) HYPOTHETICAL: NO

        (vi) ORIGINAL SOURCE:
              AB1388

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ARGTCGAAGA ASGGSGCGTA CTGSCC   26

(2) INFORMATION FOR SEQ ID NO:25:

         (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 23 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (synthetic)

       (iii) HYPOTHETICAL: NO

        (vi) ORIGINAL SOURCE:
              AB1295

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ACSARSGGYT CYTGYTCSGC YTG   23

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 36 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (synthetic)

      (iii) HYPOTHETICAL: NO

       (vi) ORIGINAL SOURCE:
            AB1296

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CTTCGTGTCC ACSGTSSWSG TSSWGATCGT GTCGAA   36

(2) INFORMATION FOR SEQ ID NO:27:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 37 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (synthetic)

      (iii) HYPOTHETICAL: NO

       (vi) ORIGINAL SOURCE:
            AB1297

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TGATGCACTC GTCSGTSGTG AASAGGTCGC AGAASGG   37

(2) INFORMATION FOR SEQ ID NO:28:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 56 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: DNA (synthetic)

      (iii) HYPOTHETICAL: NO

       (vi) ORIGINAL SOURCE:
            AB1025

       (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CGGAACTCCT GGCTGAACTG CTTCTCCTGG GTGCTCTGGG GGATGGCGGC GCCGTA   56

11

(2) INFORMATION FOR SEQ ID NO:29:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 29 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

    (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
        AB1026

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CGGAAYTCCT GVSWGAACTG CTTYTCCTG   29

(2) INFORMATION FOR SEQ ID NO:30:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 21 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

    (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
        AB1027

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CTGSGGGRATN GCNCGRCCGT A   21

(2) INFORMATION FOR SEQ ID NO:31:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 6756 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (iii) HYPOTHETICAL: NO

    (iv) ANTI-SENSE: NO

    (vi) ORIGINAL SOURCE:

(A) ORGANISM: Aspergillus ficuum (Aspergillus niger)
(B) STRAIN: NRRL 3135

(vii) IMMEDIATE SOURCE:
        (A) LIBRARY: lambda AF
        (B) CLONE: pAF2-3, pAF2-6, pAF2-7

(ix) FEATURE:
        (A) NAME/KEY: exon
        (B) LOCATION: 210..253

(ix) FEATURE:
        (A) NAME/KEY: intron
        (B) LOCATION: 254..355

(ix) FEATURE:
        (A) NAME/KEY: exon
        (B) LOCATION: 356..1715

(ix) FEATURE:
        (A) NAME/KEY: CDS
        (B) LOCATION: join(210..253, 356..1715)
        (D) OTHER INFORMATION: /codon_start= 210
              /product= "Phytase"

(ix) FEATURE:
        (A) NAME/KEY: sig_peptide
        (B) LOCATION: 210..380

(ix) FEATURE:
        (A) NAME/KEY: mat_peptide
        (B) LOCATION: 381..1712
        (C) IDENTIFICATION METHOD: experimental
        (D) OTHER INFORMATION: /function= "inositol phosphate
              phosphatase"
              /product= "Phytase"
              /evidence= EXPERIMENTAL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```
GTCGACTTCC CGTCCTATTC GGCCTCGTCC GCTGAAGATC CATCCCACCA TTGCACGTGG        60

GCCACCTTTG TGAGCTTCTA ACCTGAACTG GTAGAGTATC ACACACCATG CCAAGGTGGG       120

ATGAAGGGGT TATATGAGAC CGTCCGGTCC GGCGCGATGG CCGTAGCTGC CACTCGCTGC       180

TGTGCAAGAA ATTACTTCTC ATAGGCATC ATG GGC GTC TCT GCT GTT CTA CTT        233
                                 Met Gly Val Ser Ala Val Leu Leu
                                 -23             -20

CCT TTG TAT CTC CTG TCT  GG  GTATGCTAAG CACCACAATC AAAGTCTAAT           283
Pro Leu Tyr Leu Leu Ser  Gly
-15                 -10

AAGGACCCTC CCTTCCGAGG GCCCCTGAAG CTCGGACTGT GTGGGACTAC TGATCGCTGA       343
```

```
CTATCTGTGC AG A GTC ACC TCC GGA CTG GCA GTC CCC GCC TCG AGA AAT        392
               Val Thr Ser Gly Leu Ala Val Pro Ala Ser Arg Asn
               -8          -5                  1

CAA TCC AGT TGC GAT ACG GTC GAT CAG GGG TAT CAA TGC TTC TCC GAG        440
Gln Ser Ser Cys Asp Thr Val Asp Gln Gly Tyr Gln Cys Phe Ser Glu
5               10                  15                  20

ACT TCG CAT CTT TGG GGT CAA TAC GCA CCG TTC TTC TCT CTG GCA AAC        488
Thr Ser His Leu Trp Gly Gln Tyr Ala Pro Phe Phe Ser Leu Ala Asn
                25                  30                  35

GAA TCG GTC ATC TCC CCT GAG GTG CCC GCC GGA TGC AGA GTC ACT TTC        536
Glu Ser Val Ile Ser Pro Glu Val Pro Ala Gly Cys Arg Val Thr Phe
                40                  45                  50

GCT CAG GTC CTC TCC CGT CAT GGA GCG CGG TAT CCG ACC GAC TCC AAG        584
Ala Gln Val Leu Ser Arg His Gly Ala Arg Tyr Pro Thr Asp Ser Lys
            55                  60                  65

GGC AAG AAA TAC TCC GCT CTC ATT GAG GAG ATC CAG CAG AAC GCG ACC        632
Gly Lys Lys Tyr Ser Ala Leu Ile Glu Glu Ile Gln Gln Asn Ala Thr
    70                  75                  80

ACC TTT GAC GGA AAA TAT GCC TTC CTG AAG ACA TAC AAC TAC AGC TTG        680
Thr Phe Asp Gly Lys Tyr Ala Phe Leu Lys Thr Tyr Asn Tyr Ser Leu
85                  90                  95                  100

GGT GCA GAT GAC CTG ACT CCC TTC GGA GAA CAG GAG CTA GTC AAC TCC        728
Gly Ala Asp Asp Leu Thr Pro Phe Gly Glu Gln Glu Leu Val Asn Ser
                105                 110                 115

GGC ATC AAG TTC TAC CAG CGG TAC GAA TCG CTC ACA AGG AAC ATC GTT        776
Gly Ile Lys Phe Tyr Gln Arg Tyr Glu Ser Leu Thr Arg Asn Ile Val
            120                 125                 130

CCA TTC ATC CGA TCC TCT GGC TCC AGC CGC GTG ATC GCC TCC GGC AAG        824
Pro Phe Ile Arg Ser Ser Gly Ser Ser Arg Val Ile Ala Ser Gly Lys
        135                 140                 145

AAA TTC ATC GAG GGC TTC CAG AGC ACC AAG CTG AAG GAT CCT CGT GCC        872
Lys Phe Ile Glu Gly Phe Gln Ser Thr Lys Leu Lys Asp Pro Arg Ala
    150                 155                 160

CAG CCC GGC CAA TCG TCG CCC AAG ATC GAC GTG GTC ATT TCC GAG GCC        920
Gln Pro Gly Gln Ser Ser Pro Lys Ile Asp Val Val Ile Ser Glu Ala
165                 170                 175                 180

AGC TCA TCC AAC AAC ACT CTC GAC CCA GGC ACC TGC ACT GTC TTC GAA        968
Ser Ser Ser Asn Asn Thr Leu Asp Pro Gly Thr Cys Thr Val Phe Glu
                185                 190                 195

GAC AGC GAA TTG GCC GAT ACC GTC GAA GCC AAT TTC ACC GCC ACG TTC        1016
Asp Ser Glu Leu Ala Asp Thr Val Glu Ala Asn Phe Thr Ala Thr Phe
                200                 205                 210

GTC CCC TCC ATT CGT CAA CGT CTG GAG AAC GAC CTG TCC GGT GTG ACT        1064
```

14

```
Val Pro Ser Ile Arg Gln Arg Leu Glu Asn Asp Leu Ser Gly Val Thr
        215             220             225

CTC ACA GAC ACA GAA GTG ACC TAC CTC ATG GAC ATG TGC TCC TTC GAC      1112
Leu Thr Asp Thr Glu Val Thr Tyr Leu Met Asp Met Cys Ser Phe Asp
        230             235             240

ACC ATC TCC ACC AGC ACC GTC GAC ACC AAG CTG TCC CCC TTC TGT GAC      1160
Thr Ile Ser Thr Ser Thr Val Asp Thr Lys Leu Ser Pro Phe Cys Asp
245             250             255             260

CTG TTC ACC CAT GAC GAA TGG ATC AAC TAC GAC TAC CTC CAG TCC TTG      1208
Leu Phe Thr His Asp Glu Trp Ile Asn Tyr Asp Tyr Leu Gln Ser Leu
            265             270             275

AAA AAG TAT TAC GGC CAT GGT GCA GGT AAC CCG CTC GGC CCG ACC CAG      1256
Lys Lys Tyr Tyr Gly His Gly Ala Gly Asn Pro Leu Gly Pro Thr Gln
            280             285             290

GGC GTC GGC TAC GCT AAC GAG CTC ATC GCC CGT CTG ACC CAC TCG CCT      1304
Gly Val Gly Tyr Ala Asn Glu Leu Ile Ala Arg Leu Thr His Ser Pro
            295             300             305

GTC CAC GAT GAC ACC AGT TCC AAC CAC ACT TTG GAC TCG AGC CCG GCT      1352
Val His Asp Asp Thr Ser Ser Asn His Thr Leu Asp Ser Ser Pro Ala
        310             315             320

ACC TTT CCG CTC AAC TCT ACT CTC TAC GCG GAC TTT TCG CAT GAC AAC      1400
Thr Phe Pro Leu Asn Ser Thr Leu Tyr Ala Asp Phe Ser His Asp Asn
325             330             335             340

GGC ATC ATC TCC ATT CTC TTT GCT TTA GGT CTG TAC AAC GGC ACT AAG      1448
Gly Ile Ile Ser Ile Leu Phe Ala Leu Gly Leu Tyr Asn Gly Thr Lys
            345             350             355

CCG CTA TCT ACC ACG ACC GTG GAG AAT ATC ACC CAG ACA GAT GGA TTC      1496
Pro Leu Ser Thr Thr Thr Val Glu Asn Ile Thr Gln Thr Asp Gly Phe
            360             365             370

TCG TCT GCT TGG ACG GTT CCG TTT GCT TCG CGT TTG TAC GTC GAG ATG      1544
Ser Ser Ala Trp Thr Val Pro Phe Ala Ser Arg Leu Tyr Val Glu Met
            375             380             385

ATG CAG TGT CAG GCG GAG CAG GAG CCG CTG GTC CGT GTC TTG GTT AAT      1592
Met Gln Cys Gln Ala Glu Gln Glu Pro Leu Val Arg Val Leu Val Asn
        390             395             400

GAT CGC GTT GTC CCG CTG CAT GGG TGT CCG GTT GAT GCT TTG GGG AGA      1640
Asp Arg Val Val Pro Leu His Gly Cys Pro Val Asp Ala Leu Gly Arg
405             410             415             420

TGT ACC CGG GAT AGC TTT GTG AGG GGG TTG AGC TTT GCT AGA TCT GGG      1688
Cys Thr Arg Asp Ser Phe Val Arg Gly Leu Ser Phe Ala Arg Ser Gly
            425             430             435

GGT GAT TGG GCG GAG TGT TTT GCT TAGCTGAATT ACCTTGATGA ATGGTATGTA   1742
Gly Asp Trp Ala Glu Cys Phe Ala
```

```
TCACATTGCA  TATCATTAGC  ACTTCAGGTA  TGTATTATCG  AAGATGTATA  TCGAAAGGAT  1802

CAATGGTGAC  TGTCACTGGT  TATCTGAATA  TCCCTCTATA  CCTCGTCCCA  CAACCAATCA  1862

TCACCCTTTA  AACAATCACA  CTCAACGCAC  AGCGTACAAA  CGAACAAACG  CACAAAGAAT  1922

ATTTTACACT  CCTCCCCAAC  GCAATACCAA  CCGCAATTCA  TCATACCTCA  TATAAATACA  1982

ATACAATACA  ATACATCCAT  CCCTACCCTC  AAGTCCACCC  ATCCTATAAT  CAATCCCTAC  2042

TTACTTACTT  CTCCCCCTCC  CCCTCACCCT  TCCCAGAACT  CACCCCCGAA  GTAGTAATAG  2102

TAGTAGTAGA  AGAAGCAGAC  GACCTCTCCA  CCAATCTCTT  CGGCCTCTTA  TCCCCATACG  2162

CTACACAAAA  CCCCCACCCC  GTTAGCATGC  ACTCAGAAAA  TAATCAAAAA  TAACTAAGAA  2222

GGAAAAAAAA  GAAGAAGAAA  GGTTACATAC  TCCTCTCATA  CAAACTCCAA  GACGTATACA  2282

TCAAGATGGG  CAATCCCACC  ATTACTGATA  TCCATCTATG  AACCCATTCC  CATCCCACGT  2342

TAGTTGATTA  CTTTACTTAG  AAGAAGAAAA  AGGGAAGGGA  AGGGAAAGAA  GTGGATGGGA  2402

TTGAGTTAGT  GCTCACCGTC  TCGCAGCAAG  TTTATATTCT  TTTGTTTGGC  GGATATCTTT  2462

CACTGCTCCT  GCTGGACGTT  GTCACGGGGT  GGTAGTGGTT  GGCGGTGGTG  AGGGTCCATG  2522

ATCACTCTTG  GTTTGGGGGG  TTGTTGTTGT  CGTTGTTGTT  GTTGTTGGGT  GGGCATTTTC  2582

TTTTCTTCAC  TTGGGGATTA  TTATTTGGAA  TTGGTTAGTT  TGAGTGAGTG  GGTAATATTG  2642

AATGGGTGAT  TATTGGGAAT  GAAGTAGATT  TGGCTATGAA  TGGTTGATGG  GATGGAATGA  2702

ATGGATGGAT  GAATAGATGG  AGGCGGAAAA  GTCAGGTGGT  TTGAGGTTCG  GATTATTATC  2762

TTTGTGCCTG  AGGCATCACT  CTCCATCTAT  GTTGTTCTTT  CTATACCGAT  CTACCAGAGC  2822

TAAGTTGACT  GATTCTACCA  CAGTGCACAA  TAAGTATGTA  CTTATTTCAT  TTAGAGTATT  2882

TAGATTAACC  CGCTGTGCTA  TTTGCCGTAG  CTTTCCACCC  AATTTCGAAG  TTCGAAGAAT  2942

TAAAACTCAT  CCTACAGTAC  AGAATAGAAG  TAAAAGGAGA  AGAGAAAAAC  AAGATAATAC  3002

AACCAGTCCA  GGTCCATTCT  AGATCTCGAA  TGACCACCAA  ATAAGAAAGC  AACAAGCAAG  3062

TAAGCAAAGC  ATAAGTCTAA  ATGAACGCCA  ATAACTTCAT  CGCCTGCCTT  TGAAACTGAA  3122

CGCTATGCAC  GAATGGCTCG  AAATGATTCC  CTTAACTCCG  TAGTATTGAG  AGTGAGAGGA  3182

AAAGAAAAAA  AGAGACAGAA  AAGCTGACCA  TGGGAAAGAA  GCATGATCAG  TCGGGAATGG  3242

ATCTGCGGGT  TGAGATAGAT  ATGAGTTGCC  TCGCAGATCC  GGTGACAAGA  TAAGAGAATT  3302

GGGAGATGTG  ATCAGCCACT  GTAACTTCAT  CAAGCATCGA  CATTCAACGG  TCGGGTCTGC  3362

GGGTTGAGAT  GCAAGTTGAG  ATGCCACGCA  GACCCGAACA  GAGTGAGAGA  TGTGAGACTT  3422
```

```
TTGAACCACT GTGACTTCAT CAAGCATCAA AACACACTCC ATGGTCAATC GGTTAGGGTG    3482

TGAGGGTTGA TATGCCAGGT TCGATGCCAC GCAGACCCGA ACCGACTGAG AAATATGAAA    3542

AGTTGGACAG CCACTTCATC TTCATCAAGC GTAAAACCCC AATCAATGGT AAATCGAAAA    3602

CGAATCTGCG GGCTGATGTG GAAATGAGAC GAATGCCTCG CAGATTCGAA GACACGTAAA    3662

TCGAGATGAA CAATCACTTT AACTTCATCA AAGCCTTAAA TCACCCAATG GCCAGTCTAT    3722

TCGGGTCTGC GGGTTGAGGT TCCTGTTGAG ATGCCACGCA GACTGCGAAC ATGCGATGCA    3782

TTATAAGTTG GACGAGTGTA GACTGACCAT TGATAACCGA GATAAACAAT CACTTCAACT    3842

TCATCAAAGC CTTAAATCAC TCAATGGCCA GTCTGTTTGC GGTCTGCGGG CTGATACCCA    3902

AGTTGCGATG CCACGCAGAC TGCAAACATT GATCGAGAGA CGAGAAAAAC AACGCACTTT    3962

AACTTCAACA AAAGCCTTTC AATCAGTCAA TGGCCAGTCT GTTCGCGGTC TGCGGGCTGA    4022

TATGCGAGTT GAGGTGCCTC GCAGACCGCG AACATGCGAT GTAATTTCTT AGTTAGACGA    4082

GTGCCTGGCC ATTGAGAAAC GAGAGAAACA ACCACTTTAA CTTCATGAAA GCCTTGAACT    4142

ACTCAATGAC CCGTCTGTTG GCGGTCTGCG GGCTGATATT CGAGTTGAGA TGCCACGCAG    4202

ACCGCCAACA TGCGATGTAT CATGTAAGTT AGATGAGTGA CTGGCCATTG AGAAACGAGA    4262

GAAACAACCA CACTTCATGA GAGCCTTAAA TTATTCAATG ACCAGTCTGT TCACGGTCTG    4322

CGGGTTGGTA TGCGAGTCGA GGTGCCTCGC AGACCGCGAA CATGCGATGT TTTCGATGGA    4382

CGAGTGAAGC CTGACGATCG AGAACTATCT CAGTTGGGTT GGCCATTCGG CTGGCCGTTG    4442

GGTTTAGTAT TAGGATCGTC AGGTTTGTCC GATGGAACGT TCCGTTTGCG TGCGTTGGCG    4502

CGACGAGCCC TCTCCTCGGC GTGATTCTGA AATTCTGCAA TCAGGGCAGC CGCAGCACGG    4562

CGACGGGACG TCCTCCAGGA GCTGTGTTGA AGTTTCGGGG TGGCGGTCCA GAAGGGGGAG    4622

TTACATTAAA AGCCTCATAG ATGTCTTTGG GTGGTTCCGG GGGGCCCATC GCAAGATCTT    4682

CTGGAGTTGT GCGTCTGATC ATCTCTTGAG TGTAATTGCG ACGCAGACCG AGCTTCAGGA    4742

TTTTGGAAGG GCTGGATCGC TCCTGCTGAC TCTTTCCCTC AGCGGGCTTC GTCTCGGCAG    4802

TCTTCATTTC GGCGGGCTGA TCTTCCATCT CAGAATGGGA TCGCTTTCTG GTCGCTGCAC    4862

CCGCTCCTCC CTTCAAGGTC AGCTTGATGC GCAGCGTCTT GGGCGGCTCA GCTGGTGGAG    4922

TTGGTTCCGG CTCTGGCTCC CTCCGGCGTC GCTTGGGCAC TTGAGTAGTC TCTGAGGCTT    4982

CGCCGCGGCG CCGTTTGCGA GTCGGCTCCT TGGTCTCTTT GGCCTCTTTC ACTTCACCTG    5042

GACCGTCTTT CGGGGCGGTT TCATCGTGCT GAGCGATCAA GGTTTGGATG TAGGCAGCCG    5102
```

17

```
GCATCATTCG ATCAACGGCA ATTCCTCTCT TGCGGGCCTC CTCCCGAGCC TTGATTGTCG    5162

CCTTGACCTC GTCCACGTTT TCGAAGAAGA AAGGCATCTT GTTATCCTGA GGCAAGTTGC    5222

GCTCTCCCAT GCGTGGGGAT ATCCGAAGAT GCGGTCCTTC TCGAACTGTT CATGAGACTT    5282

CAGACGAATT GGAGGCTGGG GGAGCAATTT GTCTCCGTAG GTGTTGTTAG GGCGGAACCA    5342

AGAATAGCCT TCGCCTACAA CGACAAGCTC TTCGCCAAAT TTATTTTTTT GGCCTGTAAA    5402

AACGAACCCA TCCTCGTCAG TCCACCGGTG CGTCTCGGAC GTAGAGATTG GCTTACTTAT    5462

TCCCTCAACG CCGATCTCTG CCTGGGGCTG CGCTTCGGAT GCGGCCTCGG TCACGGCTCC    5522

GCCTCGGACT GCACCGCTGG AGTTTCGGTC TTCTTCTCCT GCTTCTCCAG GTACTCCTTG    5582

CGTAACTCTT CGATCAGCCT CGGCTTCCGA TGACTGCTCA AATTCTGGAG CAACAGCTGC    5642

CGCGGCCAGG TCAAGCAGGC GGTTTGCTAA AACTGCCCAT TTTCCATCGA CACCTGCCTC    5702

CGACGCCTGT GCAAAACCAG CTGTTTTCGC ATTGGCCTGT TTGTTGGCAC GCGTCTTCTT    5762

GACTGCTGCC TTGCCCTTTA CTTCCTTGAG AGCAGACTCT GGCTTAGATG ATGGTGCACG    5822

GTTTCTGCGG AAGCGCCGCT CAGATTCCAA AGATTCCATA GCTTTAATGG TAGGCTTTCT    5882

GGTTCTTCCA GAAGTGCGCG CAGCTGACGT AGTGGTTGAG TAGCTGGCAG TTGGGGATCC    5942

TGGGCCCTCA TTGGAACCAT CAAGACCAAA TTTGTTTCCA TACATATCAG CATGGTATTC    6002

AAAAGGAAAA CTTTCGCCGT ACGGAGTACT GCGTTCGATT CCGGGTGTAT CCAAGTCGTA    6062

TCCAGACATG GTGTCGAATT CAGCCTTGCT GTCAAGAGCA GGGGTACTTT CAATGCTGTC    6122

AGCAACCACG CGGCCAAAGG GCGTCTTCGG GAAAGAAGGT GTTTCAAGAG AAGCGTCATC    6182

CACGGCCTGG CTTGCGGCGT TGATTGCAGA CTTTCGAGTA GATCGCTGAG GTCGCGAACT    6242

GGTTCGAGTA GCAACCTGTG AATTGGCAGC CTTGTGACTG CTTCGATTCA CTGCAGAGAC    6302

GGAGTAGACT GCACTGATTT GGAATTCTGA GTCGCAGCCA TTCTGGATTT GCGTTCGGCG    6362

CGACGAGATC TCGCAGTCGT GGTACGAGGA GTAGAGCGAG GCTGCGTAGC AGTGTTGCAA    6422

GCTTGGTGCT AGCCTCCTGG GCTTCAGCAG CTTCAGCAGT GGTGGCAGAC GCAGCAGAAT    6482

TAGCGGAGCT TTATCGGCTT TGCCGCTCTG AGCGTTGGGA GTAGAAGTGA GAGAAGAGGT    6542

AGAGTCCACG GAAGAAGTCT TCTCGCTGTT CTCAAAGCCG TTCAGCTTTG CTGGCATAGA    6602

CTTACGCGTC TTGCGGCTGT TGGAAGCGGA AGAGTTCATG GCGGGAGAGG AGACGTTAGA    6662

AGTAGACATG GTGGGGTTTG TTGACGGGTT TTGAGTAACA AGAGACTTGC GTCGATCTTT    6722

GAGTGTTCTT GACAGAAAGT TATGCAACGT CGAC                                6756
```

18

(2) INFORMATION FOR SEQ ID NO:32:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 467 amino acids
          (B) TYPE: amino acid
          (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: protein

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```
Met Gly Val Ser Ala Val Leu Leu Pro Leu Tyr Leu Leu Ser Gly Val
-23          -20              -15              -10

Thr Ser Gly Leu Ala Val Pro Ala Ser Arg Asn Gln Ser Ser Cys Asp
         -5               1               5

Thr Val Asp Gln Gly Tyr Gln Cys Phe Ser Glu Thr Ser His Leu Trp
10               15              20                      25

Gly Gln Tyr Ala Pro Phe Phe Ser Leu Ala Asn Glu Ser Val Ile Ser
             30               35                  40

Pro Glu Val Pro Ala Gly Cys Arg Val Thr Phe Ala Gln Val Leu Ser
             45               50                  55

Arg His Gly Ala Arg Tyr Pro Thr Asp Ser Lys Gly Lys Lys Tyr Ser
         60               65               70

Ala Leu Ile Glu Glu Ile Gln Gln Asn Ala Thr Thr Phe Asp Gly Lys
    75               80               85

Tyr Ala Phe Leu Lys Thr Tyr Asn Tyr Ser Leu Gly Ala Asp Asp Leu
90               95               100                     105

Thr Pro Phe Gly Glu Gln Glu Leu Val Asn Ser Gly Ile Lys Phe Tyr
             110              115              120

Gln Arg Tyr Glu Ser Leu Thr Arg Asn Ile Val Pro Phe Ile Arg Ser
         125              130              135

Ser Gly Ser Ser Arg Val Ile Ala Ser Gly Lys Lys Phe Ile Glu Gly
         140              145              150

Phe Gln Ser Thr Lys Leu Lys Asp Pro Arg Ala Gln Pro Gly Gln Ser
    155              160              165

Ser Pro Lys Ile Asp Val Val Ile Ser Glu Ala Ser Ser Ser Asn Asn
170              175              180                     185

Thr Leu Asp Pro Gly Thr Cys Thr Val Phe Glu Asp Ser Glu Leu Ala
             190              195              200

Asp Thr Val Glu Ala Asn Phe Thr Ala Thr Phe Val Pro Ser Ile Arg
         205              210              215

Gln Arg Leu Glu Asn Asp Leu Ser Gly Val Thr Leu Thr Asp Thr Glu
```

19

```
                     220                     225                     230
    Val Thr Tyr Leu Met Asp Met Cys Ser Phe Asp Thr Ile Ser Thr Ser
        235             240             245

    Thr Val Asp Thr Lys Leu Ser Pro Phe Cys Asp Leu Phe Thr His Asp
    250             255             260             265

    Glu Trp Ile Asn Tyr Asp Tyr Leu Gln Ser Leu Lys Lys Tyr Tyr Gly
                270             275             280

    His Gly Ala Gly Asn Pro Leu Gly Pro Thr Gln Gly Val Gly Tyr Ala
                285             290             295

    Asn Glu Leu Ile Ala Arg Leu Thr His Ser Pro Val His Asp Asp Thr
            300             305             310

    Ser Ser Asn His Thr Leu Asp Ser Ser Pro Ala Thr Phe Pro Leu Asn
        315             320             325

    Ser Thr Leu Tyr Ala Asp Phe Ser His Asp Asn Gly Ile Ile Ser Ile
    330             335             340             345

    Leu Phe Ala Leu Gly Leu Tyr Asn Gly Thr Lys Pro Leu Ser Thr Thr
                350             355             360

    Thr Val Glu Asn Ile Thr Gln Thr Asp Gly Phe Ser Ser Ala Trp Thr
                365             370             375

    Val Pro Phe Ala Ser Arg Leu Tyr Val Glu Met Met Gln Cys Gln Ala
            380             385             390

    Glu Gln Glu Pro Leu Val Arg Val Leu Val Asn Asp Arg Val Val Pro
        395             400             405

    Leu His Gly Cys Pro Val Asp Ala Leu Gly Arg Cys Thr Arg Asp Ser
    410             415             420             425

    Phe Val Arg Gly Leu Ser Phe Ala Arg Ser Gly Gly Asp Trp Ala Glu
                430             435             440

    Cys Phe Ala
```

(2) INFORMATION FOR SEQ ID NO:33:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1404 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA to mRNA

    (iii) HYPOTHETICAL: NO

    (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```
ATGGGCGTCT CTGCTGTTCT ACTTCCTTTG TATCTCCTGT CTGGAGTCAC CTCCGGACTG      60

GCAGTCCCCG CCTCGAGAAA TCAATCCAGT TGCGATACGG TCGATCAGGG GTATCAATGC     120

TTCTCCGAGA CTTCGCATCT TTGGGGTCAA TACGCACCGT TCTTCTCTCT GGCAAACGAA     180

TCGGTCATCT CCCCTGAGGT GCCCGCCGGA TGCAGAGTCA CTTTCGCTCA GGTCCTCTCC     240

CGTCATGGAG CGCGGTATCC GACCGACTCC AAGGGCAAGA AATACTCCGC TCTCATTGAG     300

GAGATCCAGC AGAACGCGAC CACCTTTGAC GGAAAATATG CCTTCCTGAA GACATACAAC     360

TACAGCTTGG GTGCAGATGA CCTGACTCCC TTCGGAGAAC AGGAGCTAGT CAACTCCGGC     420

ATCAAGTTCT ACCAGCGGTA CGAATCGCTC ACAAGGAACA TCGTTCCATT CATCCGATCC     480

TCTGGCTCCA GCCGCGTGAT CGCCTCCGGC AAGAAATTCA TCGAGGGCTT CCAGAGCACC     540

AAGCTGAAGG ATCCTCGTGC CCAGCCCGGC CAATCGTCGC CCAAGATCGA CGTGGTCATT     600

TCCGAGGCCA GCTCATCCAA CAACACTCTC GACCCAGGCA CCTGCACTGT CTTCGAAGAC     660

AGCGAATTGG CCGATACCGT CGAAGCCAAT TTCACCGCCA CGTTCGTCCC CTCCATTCGT     720

CAACGTCTGG AGAACGACCT GTCCGGTGTG ACTCTCACAG ACACAGAAGT GACCTACCTC     780

ATGGACATGT GCTCCTTCGA CACCATCTCC ACCAGCACCG TCGACACCAA GCTGTCCCCC     840

TTCTGTGACC TGTTCACCCA TGACGAATGG ATCAACTACG ACTACCTCCA GTCCTTGAAA     900

AAGTATTACG GCCATGGTGC AGGTAACCCG CTCGGCCCGA CCCAGGGCGT CGGCTACGCT     960

AACGAGCTCA TCGCCCGTCT GACCCACTCG CCTGTCCACG ATGACACCAG TTCCAACCAC    1020

ACTTTGGACT CGAGCCCGGC TACCTTTCCG CTCAACTCTA CTCTCTACGC GGACTTTTCG    1080

CATGACAACG GCATCATCTC CATTCTCTTT GCTTTAGGTC TGTACAACGG CACTAAGCCG    1140

CTATCTACCA CGACCGTGGA GAATATCACC CAGACAGATG GATTCTCGTC TGCTTGGACG    1200

GTTCCGTTTG CTTCGCGTTT GTACGTCGAG ATGATGCAGT GTCAGGCGGA GCAGGAGCCG    1260

CTGGTCCGTG TCTTGGTTAA TGATCGCGTT GTCCCGCTGC ATGGGTGTCC GGTTGATGCT    1320

TTGGGGAGAT GTACCCGGGA TAGCTTTGTG AGGGGGTTGA GCTTTGCTAG ATCTGGGGGT    1380

GATTGGGCGG AGTGTTTTGC TTAG                                          1404
```

21

(2) INFORMATION FOR SEQ ID NO:34:

 (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 36 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

 (ii) MOLECULE TYPE: DNA (synthetic)

 (iii) HYPOTHETICAL: NO

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GGGTAGAATT CAAAAATGGG CGTCTCTGCT GTTCTA   36

(2) INFORMATION FOR SEQ ID NO:35:

 (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 33 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

 (ii) MOLECULE TYPE: DNA (synthetic)

 (iii) HYPOTHETICAL: NO

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

AGTGACGAAT TCGTGCTGGT GGAGATGGTG TCG   33

(2) INFORMATION FOR SEQ ID NO:36:

 (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 21 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

 (ii) MOLECULE TYPE: DNA (synthetic)

 (iii) HYPOTHETICAL: NO

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GAGCACCAAG CTGAAGGATC C   21

(2) INFORMATION FOR SEQ ID NO:37:

 (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 34 base pairs
  (B) TYPE: nucleic acid

```
          (C)  STRANDEDNESS: single
          (D)  TOPOLOGY: linear

     (ii)  MOLECULE TYPE: DNA (synthetic)

    (iii)  HYPOTHETICAL: NO

     (xi)  SEQUENCE DESCRIPTION: SEQ ID NO:37:

AAACTGCAGG CGTTGAGTGT GATTGTTTAA AGGG                              34


 (2)  INFORMATION FOR SEQ ID NO:38:

      (i)  SEQUENCE CHARACTERISTICS:
           (A)  LENGTH: 42 base pairs
           (B)  TYPE: nucleic acid
           (C)  STRANDEDNESS: single
           (D)  TOPOLOGY: linear

    (iii)  HYPOTHETICAL: NO

     (vi)  ORIGINAL SOURCE:
           AG-1

     (xi)  SEQUENCE DESCRIPTION: SEQ ID NO:38:

GACAATGGCT ACACCAGCAC CGCAACGGAC ATTGTTTGGC CC                     42


 (2)  INFORMATION FOR SEQ ID NO:39:

      (i)  SEQUENCE CHARACTERISTICS:
           (A)  LENGTH: 24 base pairs
           (B)  TYPE: nucleic acid
           (C)  STRANDEDNESS: single
           (D)  TOPOLOGY: linear

     (ii)  MOLECULE TYPE: DNA (synthetic)

    (iii)  HYPOTHETICAL: NO

     (vi)  ORIGINAL SOURCE:
           AG-2

     (xi)  SEQUENCE DESCRIPTION: SEQ ID NO:39:

AAGCAGCCAT TGCCCGAAGC CGAT                                         24


 (2)  INFORMATION FOR SEQ ID NO:40:

      (i)  SEQUENCE CHARACTERISTICS:
           (A)  LENGTH: 21 base pairs
           (B)  TYPE: nucleic acid
           (C)  STRANDEDNESS: single
           (D)  TOPOLOGY: linear
```

(ii) MOLECULE TYPE: DNA (synthetic)

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CTCTGCAGGA ATTCAAGCTA G   21

(2) INFORMATION FOR SEQ ID NO:41:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 36 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

    (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
        18-2

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CGAGGCGGGG ACTGCCAGTG CCAACCCTGT GCAGAC   36

(2) INFORMATION FOR SEQ ID NO:42:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 36 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

    (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
        18-3

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GTCTGCACAG GGTTGGCACT GGCAGTCCCC GCCTCG   36

(2) INFORMATION FOR SEQ ID NO:43:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 21 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single

             (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (synthetic)

        (iii) HYPOTHETICAL: NO

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GGCACGAGGA TCCTTCAGCT T   21

    (2) INFORMATION FOR SEQ ID NO:44:

        (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 12 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (synthetic)

        (iii) HYPOTHETICAL: NO

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

AATTCAAGCT TG   12

    (2) INFORMATION FOR SEQ ID NO:45:

        (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 36 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (synthetic)

        (iii) HYPOTHETICAL: NO

        (vi) ORIGINAL SOURCE:
             24-2

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CGAGCCGGGG ACTGCCAGGC GCTTGGAAAT CACATT   36

    (2) INFORMATION FOR SEQ ID NO:46:

        (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 36 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (synthetic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
         24-3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

AATGTGATTT CCAAGCGCCT GGCAGTCCCC GCCTCG   36

(2) INFORMATION FOR SEQ ID NO:47:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 36 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

    (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
         fyt-2

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

AACAGCAGAG ACGCCCATTG CTGAGGTGTA ATGATG   36

(2) INFORMATION FOR SEQ ID NO:48:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 36 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (synthetic)

    (iii) HYPOTHETICAL: NO

    (vi) ORIGINAL SOURCE:
         fyt-3

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CATCATTACA CCTCAGCAAT GGGCGTCTCT GCTGTT   36

(2) INFORMATION FOR SEQ ID NO:49:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 14 base pairs
        (B) TYPE: nucleic acid

             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (synthetic)

     (iii) HYPOTHETICAL: NO

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

AGCTTCCCCG GTAC   14

(2) INFORMATION FOR SEQ ID NO:50:

       (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 14 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (synthetic)

     (iii) HYPOTHETICAL: NO

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

AGCTCCCCCG GATC   14

(2) INFORMATION FOR SEQ ID NO:51:

       (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 10 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (synthetic)

     (iii) HYPOTHETICAL: NO

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

AGCTAGGGGG   10

(2) INFORMATION FOR SEQ ID NO:52:

       (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 10 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: DNA (synthetic)

     (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

TCGACCCCCT

28